

Probing quantum capacitance in a 3D topological insulator

D. A. Kozlov^{1,2,3}, D. Bauer³, J. Ziegler³, R. Fischer³, M. L. Savchenko^{1,2},
Z. D. Kvon^{1,2}, N. N. Mikhailov¹, S. A. Dvoretzky¹, D. Weiss³

¹*A. V. Rzhanov Institute of Semiconductor Physics, Novosibirsk 630090, Russia*

²*Novosibirsk State University, Novosibirsk, 630090, Russia and*

³*Experimental and Applied Physics, University of Regensburg, D-93040 Regensburg, Germany*

(Dated: November 3, 2015)

We measure the quantum capacitance and probe thus directly the electronic density of states of the high mobility, Dirac type of two-dimensional electron system, which forms on the surface of strained HgTe. Here we show that observed magneto-capacitance oscillations probe in contrast to magnetotransport - primarily the top surface. Capacitance measurements constitute thus a powerful tool to probe only one topological surface and to reconstruct its Landau level spectrum for different positions of the Fermi energy.

PACS numbers: 73.25.+i, 73.20.At, 73.43.-f

Three dimensional topological insulators (3D TI) represent a new class of materials with insulating bulk and conducting two-dimensional surface states [1–4]. The properties of these surface states are of particular interest as they have a spin degenerate, linear Dirac like dispersion with spins locked to their electrons k -vector [4, 5]. Strained epilayers of HgTe, examined here, constitute a 3D TI with high electron mobilities allowing the observation of Landau quantization and quantum Hall steps down to low magnetic fields [6, 7]. While unstrained HgTe is a zero gap semiconductor with inverted band structure [8, 9], the degenerate Γ_8 states split and a gap opens at the Fermi energy E_F if strained. This system is a strong topological insulator [10], explored so far by transport [6, 7, 11], ARPES [12], photoconductivity and magneto-optical experiments [13–16]; also the proximity effect at superconductor/HgTe has been investigated [17]. Since these two-dimensional electron states (2DES) have high electron mobilities of several $10^5 \text{ cm}^2/\text{V}\cdot\text{s}$, pronounced Shubnikov- de Haas (SdH) oscillations of the resistivity and quantized Hall plateaus commence in quantizing perpendicular magnetic fields [6, 7, 11], stemming from both, top and bottom 2DES. The origin of the oscillatory resistivity is Landau quantization which strongly modifies the density of states (DoS). Capacitance spectroscopy allows to directly probe the thermodynamic DoS $dn/d\mu$ (n = carrier density, μ = electrochemical potential) denoted as D , of a 3D TI. The total capacitance measured between a metallic top gate and a two-dimensional electron system (2DES) depends, besides the geometric capacitance, also on the quantum capacitance $e^2 D$, connected in series and reflecting the finite density of states D of the surface electrons [18–20]; e is the elementary charge. Below, the quantum capacitance of the top surface layer is denoted as $e^2 D_t$, the one of the bottom layer by $e^2 D_b$.

The experiments are carried out on strained 80 nm thick HgTe films, grown by molecular beam epitaxy on

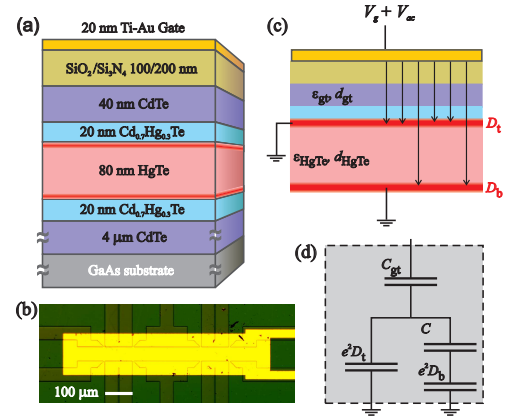


FIG. 1. (a) Schematic cross section of the heterostructures studied. The Dirac surface states, shown in red, enclose the strained HgTe layer. (b) Top view of the device showing the Hall geometry with 8 potential probes at the side, covered by the top gate (yellow). (c) Schematic illustration of the “three plate capacitor” formed by the metallic top gate electrode and top and bottom layer (red) of the two-dimensional Dirac surface electrons with density of states D_t and D_b of top and bottom layer, respectively. The geometric capacitance $C_{tb} = \epsilon_{\text{HgTe}} \epsilon_0 A / d_{\text{HgTe}}$ is determined by the gate area A , dielectric constant and thickness of CdTe, SiO₂ and Si₃N₄ layers. The electric field (black arrows) is partially screened by the top surface layer. (d) Equivalent circuit of the three plate capacitor with the quantum capacitances $Ae^2 D_t$ and $Ae^2 D_b$ in series with the respective geometrical capacitances. The equivalent circuit is the one introduced in [18], but extended by the quantum capacitance of the bottom surface, $Ae^2 D_b$. A similar equivalent circuit was recently introduced in [21].

CdTe (013) [16]. The Dirac surface electrons have high electron mobilities of order $4 \times 10^5 \text{ cm}^2/\text{V}\cdot\text{s}$. The cross-section of the structure is sketched in Fig. 1a. For transport and capacitance measurements, carried out on one and the same device, the films were patterned into Hall

bars with metallic top gates (Fig. 1b). For gating, two types of dielectric layers were used, giving similar results: 100 nm SiO₂ and 200 nm of Si₃N₄ grown by chemical vapor deposition or 80 nm Al₂O₃ grown by atomic layer deposition. In both cases, TiAu was deposited as a metallic gate. The measurements were performed at temperature $T = 1.5$ K and in magnetic fields B up to 13 T. Several devices from the same wafer have been studied. For magnetotransport measurements standard lock-in technique has been applied with the excitation AC current of 10-100 nA and frequencies from 0.5 to 12 Hz. For the capacitance measurements we superimpose the dc bias voltage V_g and a small ac bias voltage (see Fig. 1c) and measure the ac current flowing across our device phase sensitive using lock-in technique. The typical ac voltage was 50 mV at a frequency of 213 Hz. The absence of both leakage currents and resistive effects in the capacitance were controlled by the real part of the measured ac current. In order to avoid resistive effects in high magnetic fields additional measurements at lower frequencies (up to several Hz) were performed.

When the Fermi level (electrochemical potential) is located in the bulk gap the system can be viewed as a three-plate capacitor where the top and bottom surface electrons form the two lower plates (see Fig. 1c and

Typical ρ_{xx} and ρ_{xy} traces as function of the gate voltage V_g are shown in Fig. 2a. ρ_{xx} displays a maximum near $V_g = 1.5$ V, whereas ρ_{xy} changes sign; this occurs in the immediate vicinity of the charge neutrality point (CNP)[7]. The corresponding capacitance $C(V_g)$ at $B = 0$ in Fig. 2b exhibits a broad minimum between 2.2 V and 4.5 V and echoes the reduced density of states D_t and D_b of the Dirac 2DES when the Fermi energy E_F is in the gap of HgTe. For $V_g > 4.5$ V, E_F moves into the conduction band where surface electrons coexist with the bulk ones. There, the capacitance (and thus the DoS) is increased and grows only weakly with increasing V_g . Reducing V_g below 2.2 V shifts E_F below the valence band edge so that surface electrons and bulk holes coexist. A strong positive magnetoresistance, a non-linear Hall voltage and a strong temperature dependence of ρ_{xx} provide independent confirmation that E_F is in the valence band[7]. Due to the valley degeneracy of holes in HgTe and the higher effective mass, the DoS, and therefore the measured capacitance C is highest in the valence band.

For B well below 1 T both $C(V_g)$ and $\rho_{xx}(V_g)$ start to oscillate and herald the formation of Landau levels (LLs). The $C(V_g)$ trace oscillates around the zero field capacitance, shown for $B = 2$ T in Fig. 2b. These oscillations, reflecting oscillations in the DoS, are more pronounced on the electron side (to the right of the CNP). This electron-hole asymmetry stems mainly from the larger hole mass, leading to reduced LL separation on the hole side. At higher fields Hall conductivity σ_{xy} and resistivity ρ_{xy}

the corresponding equivalent circuit in Fig. 1d). From this equivalent circuit follows that, as long as D_b does not vanish, the measured total capacitance is more sensitive to changes of D_t than of D_b ; the explicit connection between D_t , D_b and the total capacitance C is given in the Supplemental Material. The ratio of $(dC/dD_t)/(dC/dD_b) = (AD_b e^2 + C_{tb})^2 / C_{tb}^2$ is significantly larger than unity since $AD_b e^2 + C_{tb}$ is (at $B = 0$) at least a factor of two larger than C_{tb} . Here, C_{tb} is the geometric capacitance $C_{tb} = \epsilon_{\text{HgTe}} \epsilon_0 A / d_{\text{HgTe}}$ between top and bottom layer, with $\epsilon_{\text{HgTe}} \approx 21$ [22], the dielectric constant of HgTe, d_{HgTe} is the thickness of the HgTe film, A is the gated TI area and ϵ_0 the dielectric constant of vacuum. Therefore, the measured capacitance reflects primarily the top surfaces DoS, D_t . In the limit $C_{tb} \rightarrow 0$ (infinite distance to bottom surface) or ($e^2 D_b \rightarrow 0$) (no charge on bottom surface) the total capacitance C is given by the expression usually used to extract the DoS D of a two-dimensional electron system: $1/C = 1/C_{\text{gt}} + 1/Ae^2 D$ with the geometric capacitance $C_{\text{gt}} = \epsilon_{\text{gt}} \epsilon_0 A / d_{\text{gt}}$, where ϵ_{gt} is the dielectric constant of the layers between gate and top 2DES, and d_{gt} , is the corresponding thickness [19, 20]. Note that $C_{\text{gt}} \ll e^2 D_t$ therefore the variations of the DoS cause only small changes of the measured value of C . (not shown) become fully quantized, for both electron and hole side. $\sigma_{xy}(V_g)$, shown for $B = 4$ T, 7 T, and 10 T in Fig. 2c, shows quantized steps of height e^2/h , ($h = \text{Planck's constant}$) as expected for spin-polarized 2DES.

An overview of transport and capacitance data in the whole V_g and B space is presented in Figs. 3d and 3e as 2D color maps (see the Supplementary Material for additional data). We start with discussing the σ_{xx} data, calculated from $\sigma_{xx} = \rho_{xx} / (\rho_{xx}^2 + \rho_{xy}^2)$, in Fig. 2d first. The sequence of σ_{xx} maxima and minima is almost symmetrical to the CNP where the electron and hole densities are equal. At fixed magnetic field the distance ΔV_g between neighboring σ_{xx} minima corresponds to a change in a carrier density Δn from which we can calculate, with the LL degeneracy eB/h , the filling rate $dn/dV_g = \alpha_{\text{total}} = 7.6 \times 10^{10} \text{ cm}^{-2}/\text{V}$ at 10 T. Comparison of electron densities extracted in the classical Drude regime with densities taken from the periodicity of SdH oscillations have shown that σ_{xx} oscillations at high B reflect the total carrier density in the TI, i.e. charge carrier densities in the bulk plus in top and bottom surfaces [7]. We therefore conclude that the filling rate α_{total} describes the change of the total carrier density n with V_g and is directly proportional to the capacitance per area $C/A = edn/dV_g$. $\alpha_{\text{total}} = 7.6 \times 10^{10} \text{ cm}^{-2}/\text{V}$ corresponds to a capacitance of $C/A = e\alpha_{\text{total}} = 1.22 \times 10^{-4} \text{ F/m}^2$, a value close to the calculated capacitance $C_{\text{gt}}^{\text{calc}}/A = 1.45 \times 10^{-4} \text{ F/m}^2$ using thickness and dielectric constant of the layers (see Figs. 1a and c and the Supplemental Material). Using this α_{total} extracted at 10 T, the Lan-

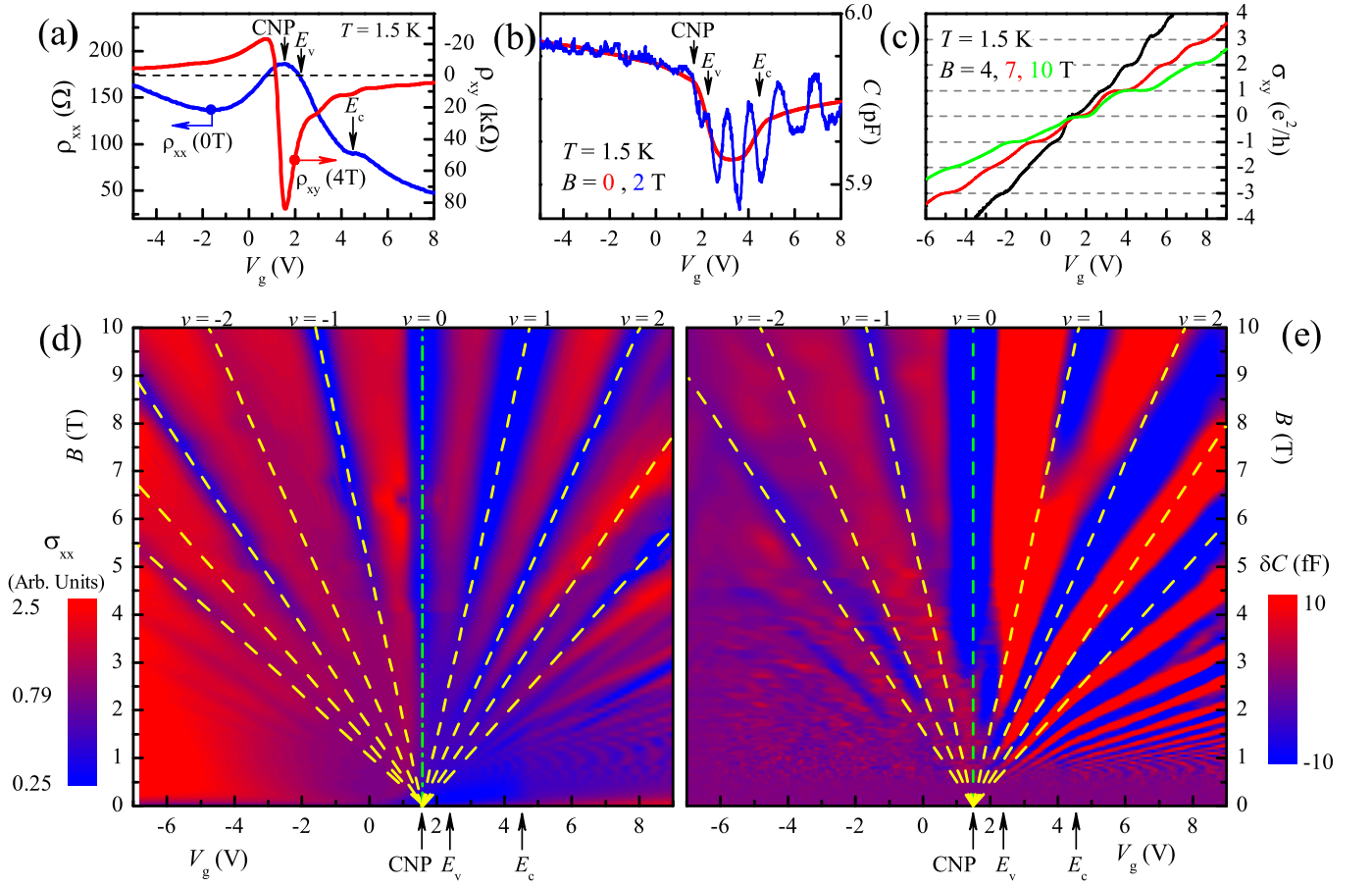


FIG. 2. (a) Typical $\rho_{xx}(V_g)$ and $\rho_{xy}(V_g)$ traces measured at $B = 0$ and $B = 4$ T, respectively. ρ_{xx} displays a maximum and ρ_{xy} changes sign in the vicinity of the charge neutrality point (CNP). (b) Capacitance measured at $B = 0$ and $B = 2$ T. The pronounced minimum of the zero field capacitance corresponds to the reduced DoS when E_F is in the gap. From that we can estimate the band edges to be at $E_v = 2.2$ V and $E_c = 4.4$ V. The quantum oscillations of the capacitance stem from the oscillations of the DoS. (c) Hall conductivity $\sigma_{xy}(V_g)$ measured for $B = 4$ T (black), 7 T (red), and 10 T (green). Quantized steps occur both on the hole side and on the electron side. (d) 2D color map of $\sigma_{xx}(V_g, B)$. For the plot each $\sigma_{xx}(V_g)$ trace has been normalized with respect to the average value, set to 1. The red color stands for σ_{xx} maxima while blue color displays minima. From the distance between blue σ_{xx} minima we extract a filling rate of $\alpha_{\text{total}} = 7.6 \times 10^{10} \text{ cm}^{-2}/\text{V}$ (see text). This allows constructing a LL fan chart (dashed yellow lines) which describes the low filling factors ν well. (e) 2D color map of $\delta C(B) = C(B) - C(0)$ shown as function of B and V_g . As in (d), the red color stands for LL maxima (DoS maxima) while blue color display the gaps between LLs. The LL fan chart is the same as in d. CNP, E_c , and E_v are marked on the x-axis.

dau level fan chart, i.e. the calculated positions of the σ_{xx} minima as function of V_g and B , fits the data for low filling factors ν quite well. For filling factors larger than $\nu = 2$ on the electron side the fan chart significantly deviates from experiment and is discussed using higher resolution data below. On the hole side where the SdH oscillations stem from bulk holes, fan chart and experimental data match almost over the whole (V_g, B) range. Remarkably, minima corresponding to odd filling factors are more pronounced than even ones, reflecting the large hole g -factor. We now turn to the magneto-capacitance data $\delta C = C(V_g, B) - C(V_g, B = 0)$ shown in Fig. 2e. The data are compared with the same fan chart derived

Previous transport experiments have shown that the periodicity of the SdH oscillations is changed at low mag-

netic fields, reflecting a reduced carrier density. Tentatively, this was ascribed to SdH oscillations stemming from transport. On the electron side, the experimental δC minima display a reduced slope compared to the calculated fan chart, pointing to a reduced filling rate. This is a first indication that the capacitance does not reflect the total carrier density in the system but predominantly the fraction of the top 2DES only. On the hole side the LL fan chart fits the data quite well but in contrast to transport, LL features are less well resolved there. This asymmetry is connected to the different effective masses; the enhanced visibility in transport is due to that fact that SdH oscillations depend on D^2 while the capacitance depends on D only.

netic fields, reflecting a reduced carrier density. Tentatively, this was ascribed to SdH oscillations stemming

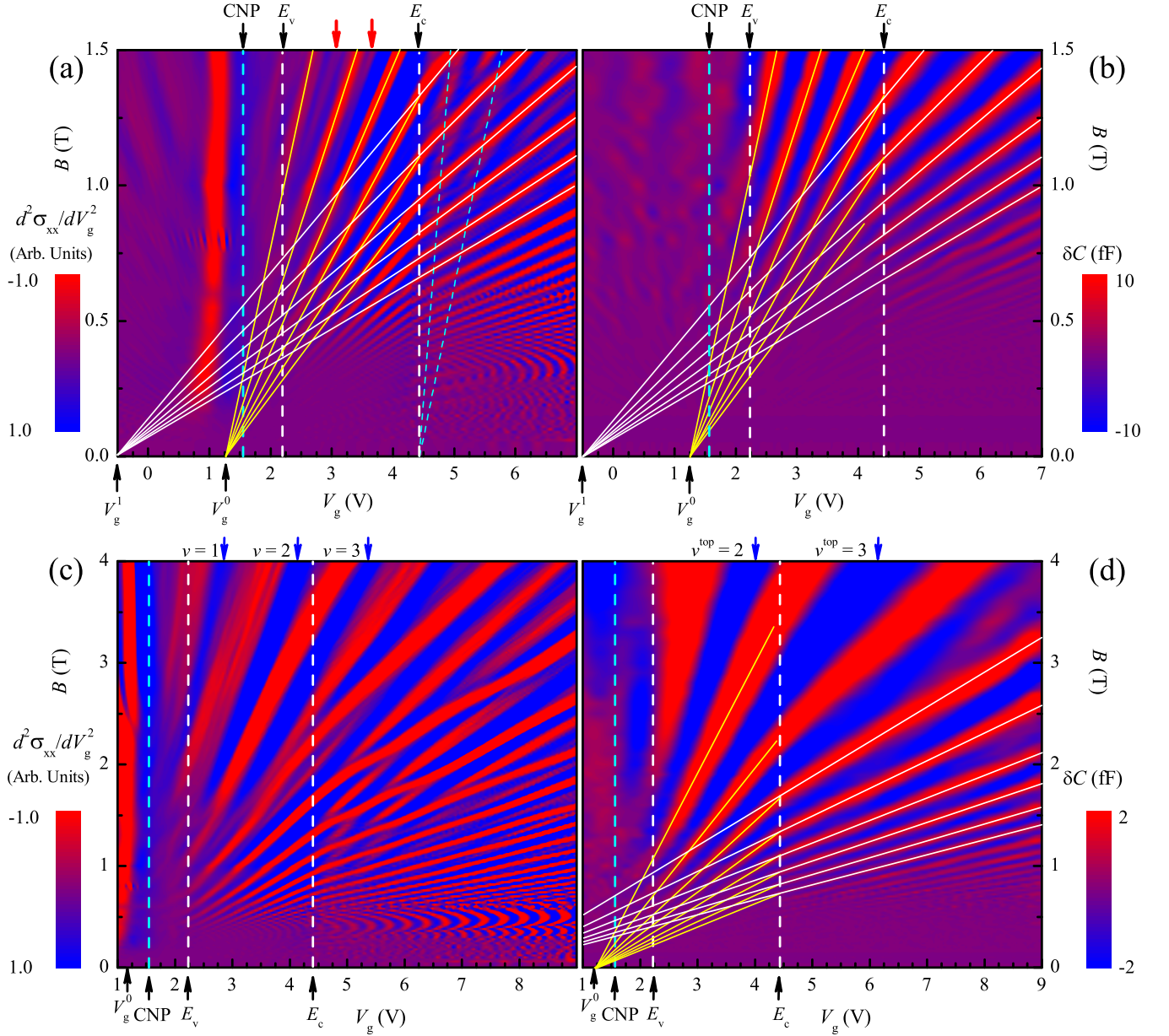


FIG. 3. (a) 2D color map of $d^2\sigma_{xx}/dV_g^2$ at low fields between 0 and 1.5 T. The second derivative brings out quantum oscillation more clearly. In each row the root mean square value was normalized to 1. Red color again indicates large σ_{xx} values, blue color low σ_{xx} , corresponding to LL gaps. The experimental data charts can be fitted by two fan charts originating at V_g^0 and V_g^1 . The different slope of the two fan charts resembles the different filling rates $\alpha_{\text{top}}^{\text{gap}}$ and $\alpha_{\text{top}}^{\text{bulk}}$ (see text). The red arrows and the two dashed lines starting at $V_g \approx 4.4$ V mark extra features discussed in the text; the gap region and CNP are marked by dashed vertical lines. (b) Corresponding $\delta C(V_g, B)$ data. The two fan charts are the same as the ones in (a). (c) As (a), but for fields between 0 and 4 T. For $V_g > 4.5$ V the data can no longer be described by simple fan charts; the pattern is entangled in a complicated way, suggesting that electrons of top and bottom surface and the bulk contribute to the conductivity oscillations. (d) Capacitance data corresponding to (c) show a quite regular Landau fan chart. The yellow and white lines in the gap and for $E_F < E_C$, respectively, correspond to the same filling rates $\alpha_{\text{top}}^{\text{gap}}$ and $\alpha_{\text{top}}^{\text{bulk}}$ as in (a) and (b). This indicates that the capacitance oscillations stem from the top surface only.

from the top surface only [7]. When top and bottom surface electrons have different electron densities and mobilities it is expected that Landau level splitting - in sufficiently low B - is in transport first resolved for the layer with the higher density and mobility, i.e. higher par-

tial conductivity and lower LL level broadening. This expectation is consistent with previous experimental observation [7]. If, indeed, the low field SdH oscillations resemble the carrier density of a single Dirac surface, capacitance oscillations which probe preferentially the

top surface and SdH oscillations should have the same period. Thus we compare σ_{xx} and capacitance δC in Fig. 3a and 3b at low B up to 1.5 T. In this low field region the capacitance (Fig. 3b) shows overall uniform oscillations of δC . The position of the maxima, corresponding to different LLs, is perfectly fitted by two fan charts featuring a distinct crossover at about $V_g = 4.4$ V. The crossover is due to E_F entering the conduction band which causes a reduced filling rate of the electrons responsible for the quantum oscillations at $V_g > 4.4$ V. Apparently, the quantum oscillations are caused by the top surface electrons only while the bulk electrons and electrons on the bottom surface merely act as a reservoir. Note, that the filling rate into the top surface state, probed in experiment, gets reduced when the filling rate into bulk states $\neq 0$, since the total filling rate must be constant [7]. From the distance of δC maxima at constant B we can extract the filling rate $\alpha_{\text{top}}^{\text{gap}}$ in the gap ($2.2 \text{ V} < V_g < 4.4 \text{ V}$) and when E_F is in the conduction band, $\alpha_{\text{top}}^{\text{bulk}}$. From Fig. 3b we obtain $\alpha_{\text{top}}^{\text{gap}} = 5.25 \times 10^{10} \text{ cm}^{-2}/\text{V}$ and $\alpha_{\text{top}}^{\text{bulk}} = 3.3 \times 10^{10} \text{ cm}^{-2}/\text{V}$. This means that in the gap $\alpha_{\text{top}}^{\text{gap}}/\alpha_{\text{total}} = 70\%$ of the total filling rate apply to the top surface while the remaining 30% can be ascribed to the bottom surface. The reduced filling rate $\alpha_{\text{top}}^{\text{bulk}}$ for E_F in the conduction band is $0.44 \cdot \alpha_{\text{total}}$ and hence the remaining filling rate of 56% is shared between bulk and back surface filling. We note that we obtain reasonable values for the filling rates only when we assume spin-resolved LL degeneracy. Since there is no signature of spin splitting down to 0.6 T, where the oscillations get no longer resolved one can conclude that the quantum oscillations always stem from non-degenerate LLs, proving the topological nature of the charge carriers. The extrapolation of the two fan charts towards $B \rightarrow 0$ defines two specific points on the V_g -axis, denoted as $V_g^0 = 1.25 \text{ V}$ and $V_g^1 = -0.5 \text{ V}$. These points correspond to vanishing electron density on the top surface n_{top} in case the respective filling rates $\alpha_{\text{top}}^{\text{gap}}$ and $\alpha_{\text{top}}^{\text{bulk}}$ would stay constant over the entire V_g range. This is not the case as $\alpha_{\text{top}}^{\text{bulk}} = \text{constant}$ only applies for E_F in the conduction band and $\alpha_{\text{top}}^{\text{gap}} = \text{constant}$ for E_F in the gap. Moving E_F into the valence band greatly reduces this filling rate. Therefore V_g^0 and V_g^1 correspond only to virtual zeroes of the electron density while the real one is much deeper in the valence band.

To get a better resolution of the low field SdH oscillation we plot $d^2\sigma_{xx}/dV_g^2$ in Fig. 3a; as before red regions indicate σ_{xx} maxima. The same LL fan chart in Figs. 3b and 3a, fits both, transport and capacitance, quite well. Hence, at sufficiently low B , both σ_{xx} and δC oscillations resemble the carrier density of the top surface only. How-

ever, even at low B striking deviations in $d^2\sigma_{xx}/dV_g^2$ pop up which get more pronounced at higher B (see below). Two faint lines at $B = 1 \dots 1.5 \text{ T}$ and $V_g = 2.5 \dots 3.5 \text{ V}$ which appear between the LLs of the top surface and are marked by red arrows in Fig. 3a, are ascribed to SdH oscillations stemming from the bottom surface. More differences occur once E_F enters the conduction band ($V_g > 4.4 \text{ V}$). New maxima appear in the fan chart and form rhomb like structures. These anomalous structures, completely absent in capacitance, get even more intriguing at higher B , displayed in Fig. 3c. Data taken at B up to 4 T, displayed in Figs. 3c and 3d, show marked differences between transport and capacitance. δC maxima, given by the red regions in Fig. 3d are, as before, well described by the same two LL fan chart with the same filling rates as in Figs. 3a and b. The filling factors given on top of Fig. 3d are the ones of the electrons in the top surface only, while the filling factors given in Fig. 3c are the ones of the total carrier density, determined by the total filling rate and the position of the σ_{xy} plateaus in Fig. 2c. The transport data for E_F in the gap show splitting of the Landau levels and for $V_g > 4.4 \text{ V}$, i.e. for E_F in the conduction band, a very complex structure with crossing LLs evolves, which is strikingly different from the one observed in the capacitance data. We thus conclude that in transport experiments the three available transport channels (top, bottom surface electrons, bulk electrons) contribute to the signal and lead to a complicated pattern of the quantum oscillations as a function of B and V_g . The oscillations of the quantum capacitance, in contrast, stem preferentially from the top surface and allow probing the LL spectrum of a single Dirac surface in a wide range of B and V_g . Corrections to that likely occur at high B and V_g : a level splitting at ($V_g \approx 7 \text{ V}$, $B \approx 3 \text{ T}$) in Fig. 3d suggest that signals from bulk or back surface can affect also δC at higher B , although to a far lesser degree when compared to transport.

In summary, we present first measurements of the quantum capacitance of a TI which directly reflect the DoS of Dirac surface states. The oscillations of the quantum capacitance in quantizing magnetic fields allow tracing the LL structure of a single Dirac surface. The complementary information provided by transport and capacitance experiments is promising in getting a better understanding of the electronic structure of TIs, the latter being particularly important for potential applications of this new class of materials.

We acknowledge funding by the Elite Network of Bavaria and by the German Science Foundation via SPP 1666. This work was partially supported by RFBR grants No 14-02-31631, 15-32-20828 and 15-52-16008.

Supplementary Information: “Probing quantum capacitance in a 3D topological insulator”

D. A. Kozlov, D. Bauer, J. Ziegler, R. Fischer, M. L. Savchenko, Z. D. Kvon, N. N. Mikhailov, S. A. Dvoretzky and D. Weiss

SAMPLES DETAILS

The experiments were carried out on strained 80 nm thick HgTe films, grown by molecular beam epitaxy (MBE) on CdTe (013) [16]. The Dirac surface electrons have high electron mobilities of order $\mu \approx 4 \times 10^5 \text{ m}^2/\text{V}\cdot\text{s}$ [7]. The cross section of a structure is sketched in Fig. S4a. For transport and capacitance measurements, carried out on one and the same device, the films were patterned into Hall bars with metallic top gates (Fig. S4b).

For gating, two types of dielectric layers were used, giving similar results: 100 nm SiO_2 and 200 nm of Si_3N_4 grown by chemical vapor deposition or 80 nm Al_2O_3 grown by atomic layer deposition (not shown in Fig. S4a). In both cases, TiAu was deposited as a metallic gate. The measurements of the total capacitance C between the gate and the HgTe layer were performed at temperature $T = 1.5 \text{ K}$ and in magnetic fields up to 13 T. Several devices from the same wafer have been studied. For magnetotransport measurements standard lock-in technique has been applied with excitation AC currents of 10-100 nA at frequencies $\omega/2\pi$ ranging from 0.5 to 12 Hz. For the capacitance measurements we superimpose the DC bias voltage with a small AC bias voltage (see Fig. 2c) and measure with lock-in technique the AC current flowing across our device phase sensitively. The typical AC voltage was 50 mV at a frequency of 213 Hz. The absence of both leakage currents and resistive effects in the capacitance were controlled by the real part of the measured AC current. These resistive effects occur when the condition $RC\omega \ll 1$ becomes invalid, for example when the series resistance R , i.e. the resistance of the two-dimensional electron gas increases at integer filling factors and high magnetic fields. In order to suppress resistive effects in high magnetic fields additional measurements at lower frequencies (down to several Hz) were performed (see below).

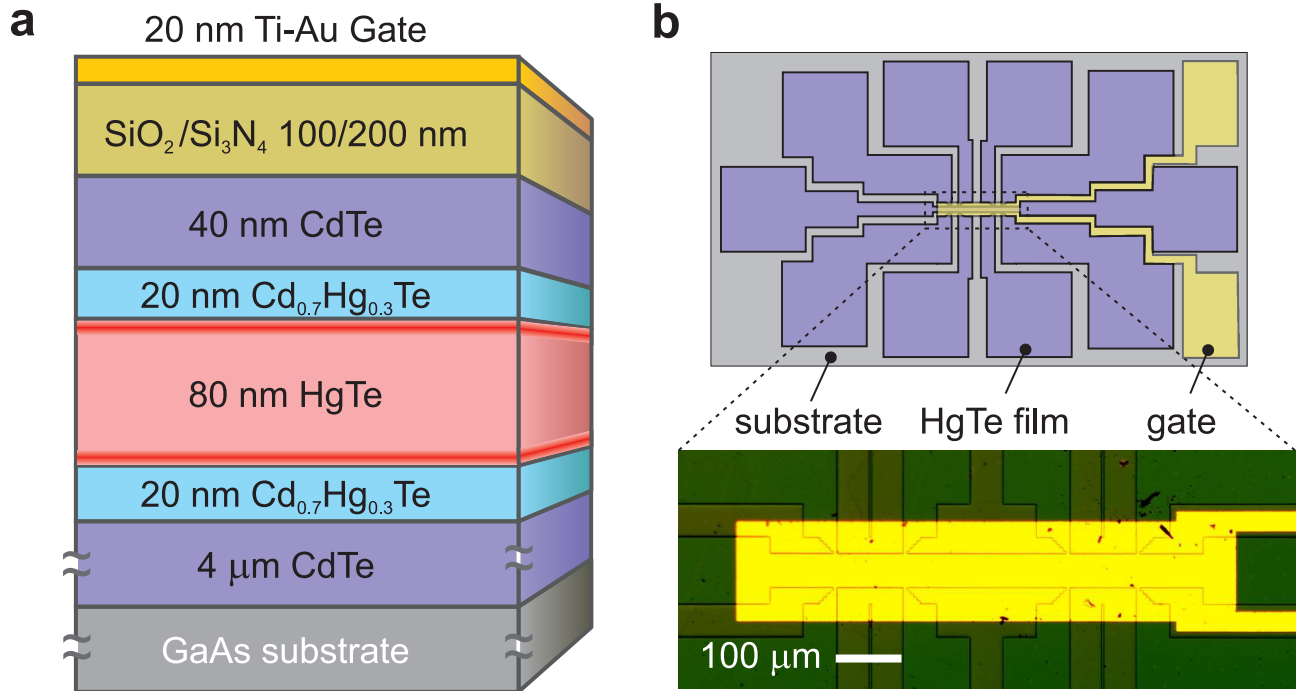


FIG. S4. **a**, Cross section of the heterostructure with $\text{SiO}_2/\text{Si}_3\text{N}_4$ insulator. The Dirac surface states, shown in red, enclose the strained HgTe layer. **b**, Top view of the device showing the Hall geometry with 10 potential probes, covered by the top gate (yellow); the gate is shown in more detail in the optical micrograph on the bottom.

SAMPLES BAND DIAGRAM AND EQUIVALENT CIRCUIT

Electrostatic and band diagram

Following the papers of Stern [23], Smith [24] and Luryi [18] we introduce an equivalent circuit for the sample's capacitance based on simple one-dimensional electrostatics. The discussion is limited to the case where the Fermi level is located in the gap, so that contributions of bulk holes and electrons are absent. The top and bottom surface layers are treated as infinitely thin, negatively charged surfaces with finite density of states (DoS), i.e. we ignored the spatial distribution of the surface electrons. Finally, we treated the insulating layer as uniform with an average dielectric constant and with the total thickness of the layers given in Fig. S4.

A cartoon of the resulting simplified band diagram is presented for zero gate voltage V_g in Fig. S5a. Without applied V_g the Fermi level is constant across the structure. The density of electrons n_t and n_b on top and bottom surface, respectively, depends on the position of the Fermi level with respect to the Dirac point (which is, other than sketched in Fig. S5a below the valence band edge). For the sake of simplicity we assume that the densities are the same at $V_g = 0$. This simplification is justified since we are only interested in changes of the carrier density with gate voltage and not in absolute values of the carrier density.

For an applied gate voltage $V_g \neq 0$ we assume that the voltage drop occurs exclusively across the $\text{SiO}_2/\text{Si}_3\text{N}_4$ insulator leading to an energy difference eV_g between the Fermi level in the gate metal and the HgTe layer. Here, e is the elementary charge. Applying a positive gate voltage $V_g > 0$ increases the carrier density in top and bottom layer by Δn_{top} and Δn_{bot} , respectively. Due to screening of the electric field by the top surface layer the induced electron density is higher in the top layer than in the bottom one; while the Fermi level throughout the HgTe layer is constant the different electron density in top and bottom surface layer causes an electrical potential drop between bottom and top surfaces of $e\phi_{\text{HgTe}}$. The change of carrier density in top and bottom layer can be written as $\Delta\mu_{\text{top(b)}} = \Delta n_{\text{top(bot)}}/D_{\text{t(b)}}$, where $D_{\text{t(b)}}$ is the density of states (DoS) of the electrons on the top (bottom) surface. Thus the following relation holds:

$$\Delta\mu_{\text{top}} = \Delta\mu_{\text{bot}} + e\phi_{\text{HgTe}}. \quad (1)$$

Using charge neutrality and Gauss's law of electrostatics the potential drop in the HgTe layer can be calculated from the electric field $E_{\text{HgTe}} = e\Delta n_b/\varepsilon_{\text{HgTe}}\varepsilon_0$, induced by the additional electron density Δn_b on the bottom layer:

$$\phi_{\text{HgTe}} = E_{\text{HgTe}}d_{\text{HgTe}} = e\Delta n_{\text{bot}}d_{\text{HgTe}}/\varepsilon_{\text{HgTe}}\varepsilon_0. \quad (2)$$

$\varepsilon_{\text{HgTe}}$ is the dielectric constant of HgTe and d_{HgTe} is the distance between the electron wave functions of top and bottom surfaces layers, approximated by the thickness of the HgTe layer, d_{HgTe} . Combining formulas 1 and 2 gives a relation between Δn_{top} and Δn_{bot} :

$$\Delta n_{\text{top}}/D_t = \Delta n_{\text{bot}}/D_b + e^2\Delta n_{\text{bot}}d_{\text{HgTe}}/\varepsilon_{\text{HgTe}}\varepsilon_0. \quad (3)$$

This is the expression given by Luryi [18] generalized for non-parabolic dispersion of the surface electrons.

The equivalent circuit

In the main text we discuss the equivalent circuit of a three plate capacitor as a model for the capacitance measured in a topological insulator with metallic gate and top and bottom surface layers. The equivalent circuit is applicable when the Fermi level is in the bulk gap of HgTe and is shown in Fig. 1d and the Supplementary Figure S5d. A similar equivalent circuit has been recently suggested for the same system but a slightly different configuration [21]. The four capacitors drawn in Supplementary Figure S5d represent either geometrical (C_{gt} and C_{tb}) or quantum (Ae^2D_t and Ae^2D_b) capacitances. Here, $C_{gt} = A\varepsilon_0\varepsilon_{\text{ins}}/d_{\text{ins}}$ is the geometrical capacitance associated with the insulating layer between the gate and the top surface of the HgTe layer; $C_{tb} = A\varepsilon_0\varepsilon_{\text{HgTe}}/d_{\text{HgTe}}$ is responsible for the potential drop across the HgTe layer between electrons on the top and bottom surface; Ae^2D_t and Ae^2D_b are the corresponding quantum capacitances representing the finite DoS of the corresponding HgTe surface states; in all cases A is the gate area. The equivalent circuit in Fig. S5d describes the electrostatics of the system: Indeed, the charge on the quantum capacitors is the one induced by the gate voltage, Δn_{top} and Δn_{bot} and the voltage across the geometrical ones is, as it should be, proportional to the corresponding induced charge and the thickness of the respective dielectric layer.

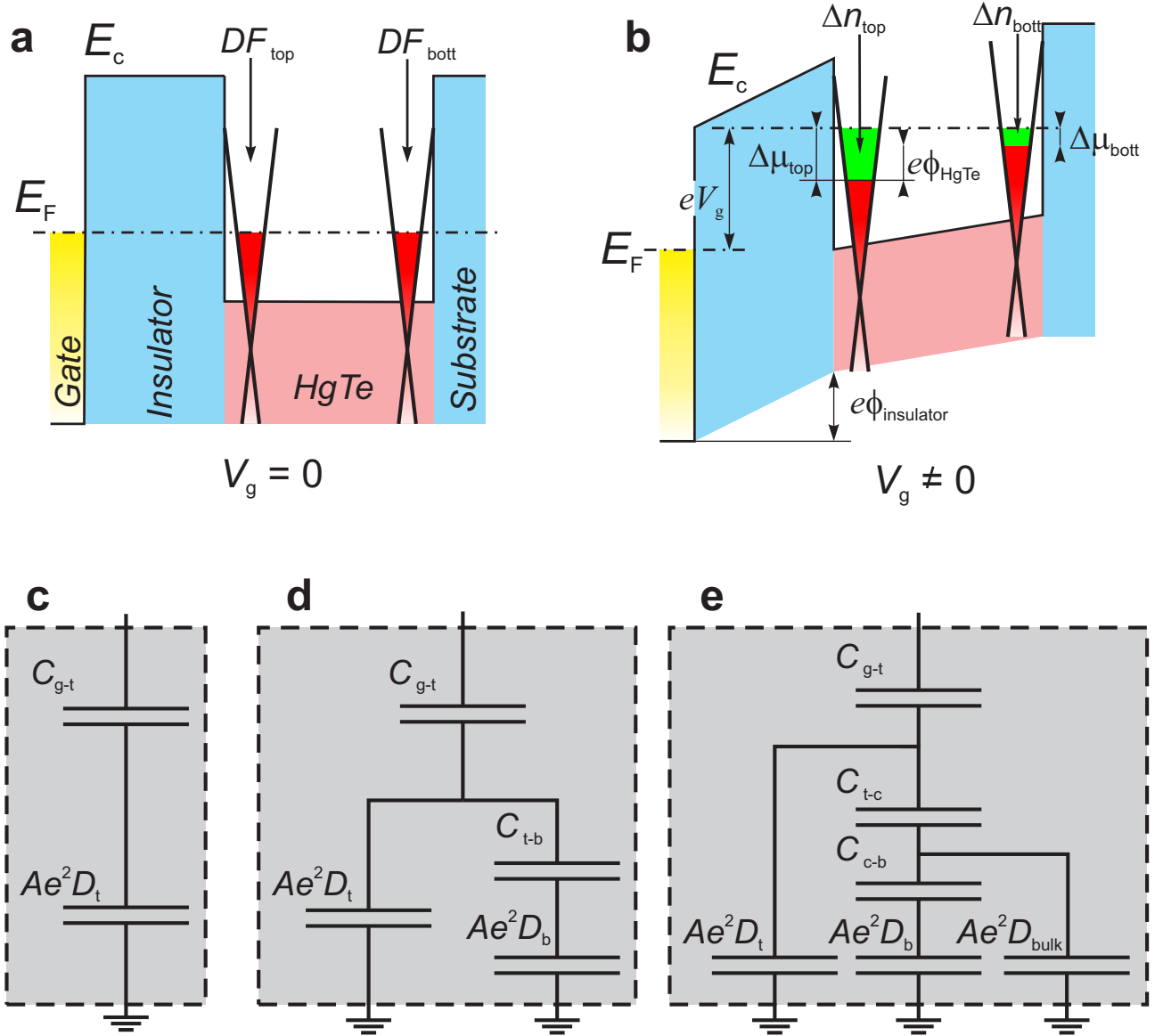


FIG. S5. **a** and **b**, Simplified band diagrams of the structure studied for E_F is in the bulk gap at $V_g = 0$ (panel a) and $V_g > 0$ (panel b). **c**, **d** and **e**, Equivalent circuit for a simple capacitor where one plate consists of a two-dimensional electron gas (c), for our three plate capacitor with E_F in the bulk gap (d) and for a three plate capacitor with E_F in the conduction band or valence band, respectively (e).

The circuit becomes simpler if one removes the influence of the bottom surface: this can be done by increasing the distance between top and bottom layer ($C_{tb} \rightarrow 0$) or by removing the charge carriers ($e^2D_b \rightarrow 0$). Then only two capacitors are present in the reduced circuit (Fig. S5c); this is the situation relevant for a gated two-dimensional electron system. In contrast, if the Fermi level is lying in the bulk electron or hole band (see Fig. S5e) the situation becomes more complex. In this case bulk carriers are characterized by a particular wave function across the HgTe layer. A simple approach to model this is to replace the capacitor C_{tb} by two capacitors C_{tc} and C_{cb} representing the potential drop between the maximum of the bulk carriers' wave function (denoted by subscript "c" at a point around the center of the HgTe layer) and the top and bottom of the HgTe layer, respectively. For this the relation $C_{tb}^{-1} = C_{tc}^{-1} + C_{cb}^{-1}$ holds.

The total sample's capacitance C for the circuit shown in Fig. S5c can be easily derived using Ohm's law and is given by:

$$\frac{A}{C} = \frac{A}{C_{\text{gt}}} + \frac{1}{e^2 D_t + \frac{1}{\frac{A}{C_{\text{tb}}} + \frac{1}{e^2 D_b}}} = AC_{\text{gt}}^{-1} + \left[e^2 D_t + \left(AC_{\text{tb}}^{-1} + (e^2 D_b)^{-1} \right)^{-1} \right]^{-1} \quad (4)$$

In the actual device the relation $C_{\text{gt}}/A \ll e^2 D_t, e^2 D_b$ holds, therefore $C \approx C_{\text{gt}}$ and any variation of the DoS, e.g. in a magnetic field, leads only to a small change of the total capacitance C . An important result which can be derived from Eq. 4 is the sensitivity of the total capacitance C to a change of the DoS of top and bottom layer, D_t and D_b , respectively:

$$dC/dD_t = \frac{Ae^2 C_{\text{gt}}^2 C_{\text{tb}}^2 (C_{\text{tb}} + Ae^2 D_b)^2}{\left[C_{\text{gt}}(C_{\text{tb}} + Ae^2 D_b) + Ae^2 (C_{\text{tb}}(D_b + D_t) + Ae^2 D_b D_t) \right]^2} \quad (5)$$

$$dC/dD_b = \frac{Ae^2 C_{\text{gt}}^2 C_{\text{tb}}^2}{\left[C_{\text{gt}}(C_{\text{tb}} + Ae^2 D_b) + Ae^2 (C_{\text{tb}}(D_b + D_t) + Ae^2 D_b D_t) \right]^2} \quad (6)$$

The ratio of both quantities

$$\frac{dC/dD_t}{dC/dD_b} = \frac{(C_{\text{tb}} + AD_b e^2)^2}{C_{\text{tb}}^2} = \left(1 + \frac{AD_b e^2}{C_{\text{tb}}} \right)^2 \quad (7)$$

is always larger than 1 as long as $D_b \neq 0$ which means that the total capacitance C is more sensitive to oscillations of the top layer, e.g., the same oscillation amplitude of D_t and D_b will lead to significantly different oscillation amplitudes in the measured total capacitance C .

Numeric evaluations

Here, we provide the values of the capacitances of our devices. The insulating layer underneath the metallic gate consists of Si_3N_4 (200 nm layer with $\varepsilon = 7.5$), SiO_2 (100 nm with $\varepsilon = 3.5$), CdTe (40 nm with $\varepsilon = 10.2$), CdHgTe (20 nm with $\varepsilon \approx 13$) and HgTe (5-7 nm with $\varepsilon \approx 21$) and $C_{\text{gt}}^{\text{calc}}/A = 1.45 \times 10^{-10} \text{ F/mm}^2$. The calculated value of the specific geometrical capacitance $C_{\text{gt}}^{\text{calc}}/A$ is very close to the experimental one, $C_{\text{gt}}/A = e\alpha_{\text{total}} = 1.22 \times 10^{-10} \text{ F/mm}^2$ obtained from the filling rate $\alpha_{\text{total}} = 7.6 \times 10^{10} \text{ cm}^{-2}/\text{V}\cdot\text{s}$ (see main text).

Using the gate area $A = 0.05 \text{ mm}^2$ (see Fig. S4b) the sample's geometric capacitance is expected to be $C_{\text{gt}} \approx 6 \text{ pF}$ which is in line with the measured value ($\approx 23 \text{ pF}$) if one subtracts the parasitic capacitance of our set up, being around 17 pF.

With the known values of the surface electrons' effective mass $m^* = 0.030m_0$ [16] one can estimate the specific quantum capacitances: $e^2 D_{\text{t(b)}}^{\text{eval}} = e^2 m^*/2\pi\hbar^2 \approx 10^{-8} \text{ F/mm}^2$ which is almost two orders of magnitude higher than the specific geometrical capacitance $C_{\text{gt}}^{\text{calc}}/A$.

Finally, the evaluated specific capacitance of the 80 nm thick HgTe layer with $\varepsilon = 21$ is $C_{\text{tb}}^{\text{calc}}/A = 2.4 \times 10^{-9} \text{ F/mm}^2$. From this value and with Eq. 3 one can derive the relation between filling rates of electrons on top and bottom surfaces: $\alpha_{\text{top}}/\alpha_{\text{bott}} \equiv (dn_{\text{top}}/dV_g)/(dn_{\text{bot}}/dV_g) = \left(\frac{D_t}{D_b} + \frac{Ae^2 D_t}{C_{\text{tb}}^{\text{eval}}} \right) \approx 5.1$. However, in our experiment, described in the main text, we obtain $\alpha_{\text{top}} = 0.7 \times \alpha_{\text{total}}$, $\alpha_{\text{bot}} = 0.3 \times \alpha_{\text{total}}$ and $\alpha_{\text{top}}/\alpha_{\text{bot}} \approx 2.3$. This discrepancy suggests that the equivalent circuit in Fig. S5d does not fully describe the experimental situation. Most likely this deviation stems from the fact that we neglected the dependence of the surface states' dispersion on the transversal electric field described in [11]. Phenomenologically, this shortcoming can be corrected by introducing a larger effective capacitance of $C_{\text{tb}}^{\text{calc}}/A = 7.5 \times 10^{-9} \text{ F/mm}^2$, indicating an enhanced screening in HgTe . This can be achieved by a higher value of the dielectric constant of HgTe or a reduced effective layer thickness.

We note that for both values of C_{tb} given above, the ratio $(dC/dD_t)/(dC/dD_b)$ is larger than 5.5, meaning that the sample capacitance is predominantly sensitive to changes of the DoS of the top surface.

DENSITIES AND PARTIAL FILLING RATES

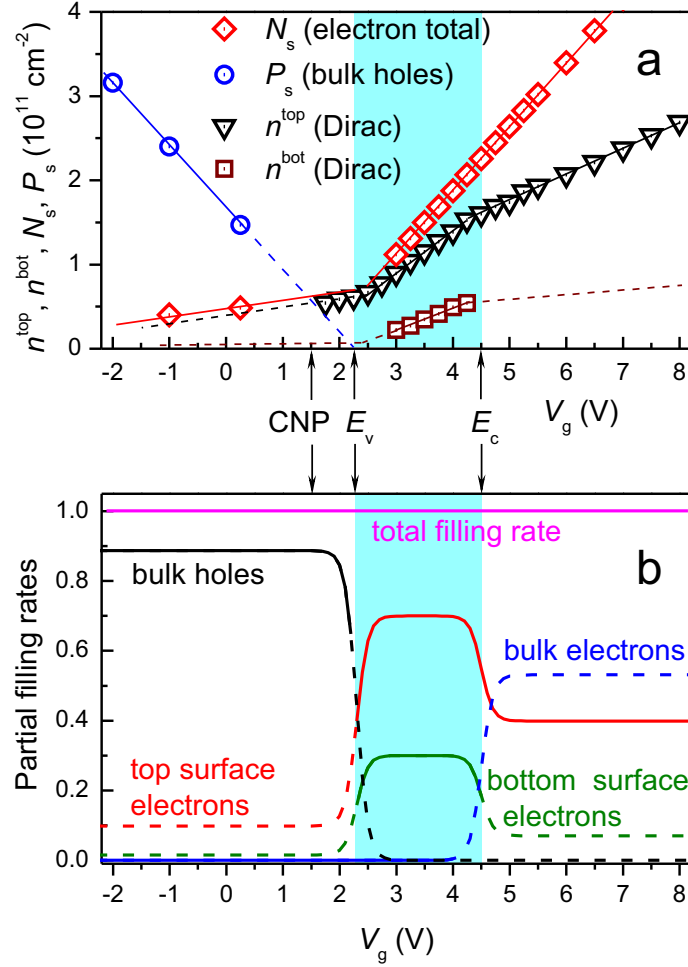


FIG. S6. **a**, Gate voltage dependence of the different charge carrier species. n_{top} is the electron density of the top surface; n_{bot} the one of the bottom surface; N_s is the total density of both surfaces plus the bulk electron density; P_s is the bulk hole density. The dashed lines extrapolate the corresponding charge carrier species in regions where data can not be directly extracted. **b**, The absolute value of the partial filling rates between surface and bulk carriers. The bluish area on both panels corresponds the bulk gap region. "CNP", " E_v " and " E_c " mark the charge neutrality point, top of the valence band and bottom of the conductive band, respectively. Values of the partial filling rates given by dashed lines can not be directly extracted from experiment but are taken from the corresponding equivalent circuit.

The gate voltage dependence of the different carrier species is shown in Fig. S6a. The total electron density N_s and the bulk hole density P_s were obtained from the low field transport data using the classical (two-carrier) Drude model as described in [7]. The top surface electron density n_{top} was obtained from the periodicity $\Delta(1/B)$ of the capacitance oscillations, $n_{\text{top}} = e(h\Delta(1/B))^{-1}$, with the period $\Delta(1/B)$ on the $1/B$ scale and Planck's constant h . For $V_g < 1.5$ this technique is not applicable since the oscillations are dominated by bulk holes. The density of the bottom surface electrons was obtained from $n_{\text{bot}} = N_s - n_{\text{top}}$ which holds as long as E_F is located in the bulk gap, i.e. $V_g = 2.2 \dots 4.5$ V. In our experiment it is impossible to directly measure the density of electrons on the bottom surface since their response is masked by electrons on the top surface. Consequently the value of n^{bot} can not be derived when the Fermi level is in the valence (because of unknown value of n^{top}) or the conductance band (because of the unknown value of n^{bulk}). However, based on the equivalent circuit shown in Fig. S5d and e one can make an educated guess on the partial filling rates. The result of such analysis is shown in Fig. S6b.

RESISTIVE EFFECTS IN THE CAPACITANCE

As mentioned in the main text, resistive effects occur when the condition $RC\omega \ll 1$ becomes invalid, for example when the resistance R of a two-dimensional electron gas connected in series to the capacitance increases at integer filling factors and high magnetic fields. The absence of both leakage currents and resistive effects in the capacitance were controlled by the real part of the measured AC current. A direct manifestation of resistive effects is the frequency-dependence of the measured $C(V_g)$ traces. A corresponding example, taken from a sample with larger gate area and therefore larger capacitance is shown in Fig. S7. The capacitance minimum appearing at the charge neutrality point (CNP) and connected to a gap opening at higher magnetic fields shows a pronounced frequency dependence. The resistive effects are weak up to 34 Hz. With increasing frequency, however, the minimum becomes deeper and the capacitance signal no longer reflects the density of states. Remarkably, in all our samples the resistive effects first appear close the CNP and are much less pronounced at integer filling factors at the same values of magnetic field. Thus the remaining capacitance oscillations reflect the density of states.

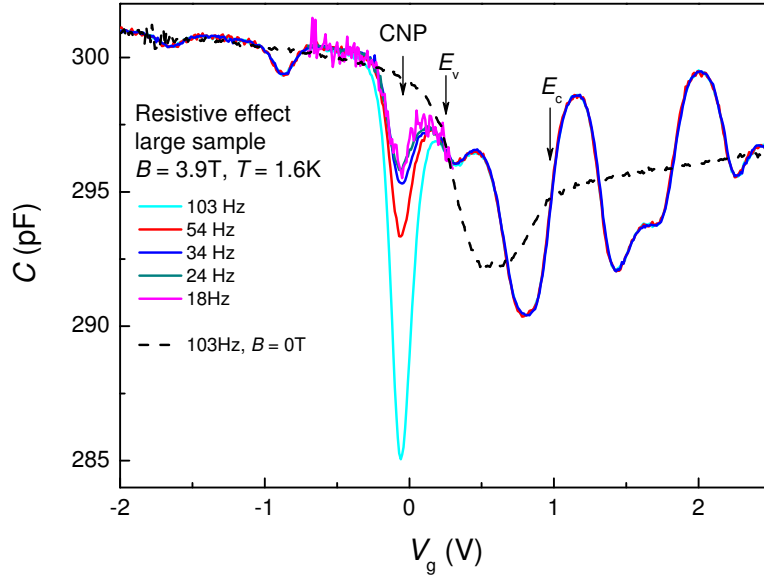


FIG. S7. Frequency dependence of the magnetocapacitance data: $C(V_g)$ measured at a magnetic field of 3.9 T and at frequencies $f = 18, 24, 34, 54$ and 103 Hz. The dashed line corresponds to the zero field $C(V_g)$ trace.

EXPERIMENTAL DATA

In this section we show some of the raw data and outline how they were prepared for the color maps in the main article. Fig. S8a displays the capacitance as a function of gate voltage V_g and for magnetic fields between 0 and 13 T. In the color maps we plot the differential magnetocapacitance $\delta C(V_g) = C(V_g)|_B - C(V_g)|_{B=0}$, shown for magnetic fields up to 2 T in Fig. S8b. In order to suppress the noise we used standard band-pass Fourier filtering with the cutoff frequencies $f_1 = 0.1 - 0.3 \text{ T}^{-1}$ and $f_2 = 1 - 5 \text{ T}^{-1}$ depending on the magnetic field range.

The corresponding transport data measured on the same device are plotted in Fig. S9a. Together with the corresponding $\rho_{xy}(V_g)$ traces (not shown) the resistivity data were converted into $\sigma_{xx}(V_g)$ traces shown in Fig. S9b. Since the value of σ_{xx} varies over several orders of magnitude and in order to improve the clarity of the color maps (Fig. 3d) each $\sigma_{xx}(V_g)$ trace was normalized with respect to its average value using $\sigma_{xx}^{\text{normalized}}(V_g) = \sigma_{xx}(V_g) / \langle \sigma_{xx} \rangle$, where $\langle \dots \rangle$ means averaging over the whole V_g range. After the normalization procedure the average value of each trace is equal to 1.

The $\sigma_{xx}(V_g)$ data taken at lower magnetic fields are shown in the Fig. S10a. To work out the oscillations more clearly we plot the second derivative $d^2\sigma_{xx}(V_g)/dV_g^2$ of the traces, shown in panel a of Fig. S10b. In order to visualize the low-field oscillations clearly on the color maps each $d^2\sigma_{xx}(V_g)/dV_g^2$ trace was normalized by its root mean square (RMS) value taken over the whole V_g scale. Finally we applied standard low-pass Fourier filtering to remove random noise which is quite pronounced for the lowest field ($B \leq 0.3 \text{ T}$) traces.

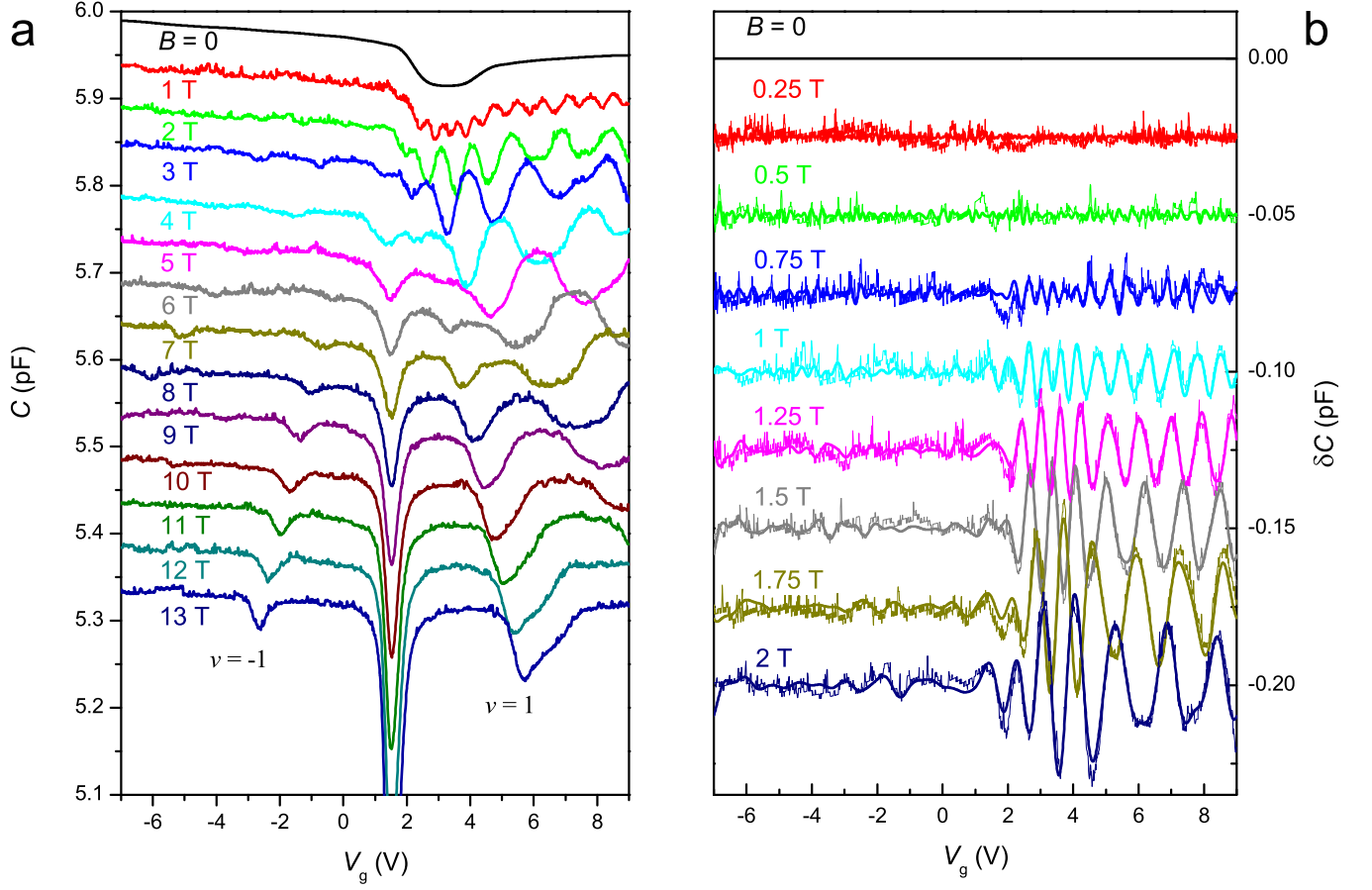


FIG. S8. Magnetocapacitance data: **a**, $C(V_g)$ dependencies at magnetic fields from 0 (on top) to 13 T (bottom) using steps of 1 T. Starting at $B = 1$ T each subsequent trace is shifted by -0.05 pF on the Y-axis. The filling factors $\nu = -1$ and 1 are indicated. **b**, Differential magnetocapacitance $\delta C(V_g) = C(V_g)|_B - C(V_g)|_{B=0}$ dependencies at magnetic fields from 0 (top) to 2 T (bottom) using steps of 0.25 T between the traces. Starting at $B = 0.25$ T each subsequent trace is shifted by -0.025 pF on the Y-axis. The thin lines correspond to the raw data and the thicker ones are the data after Fourier filtering.

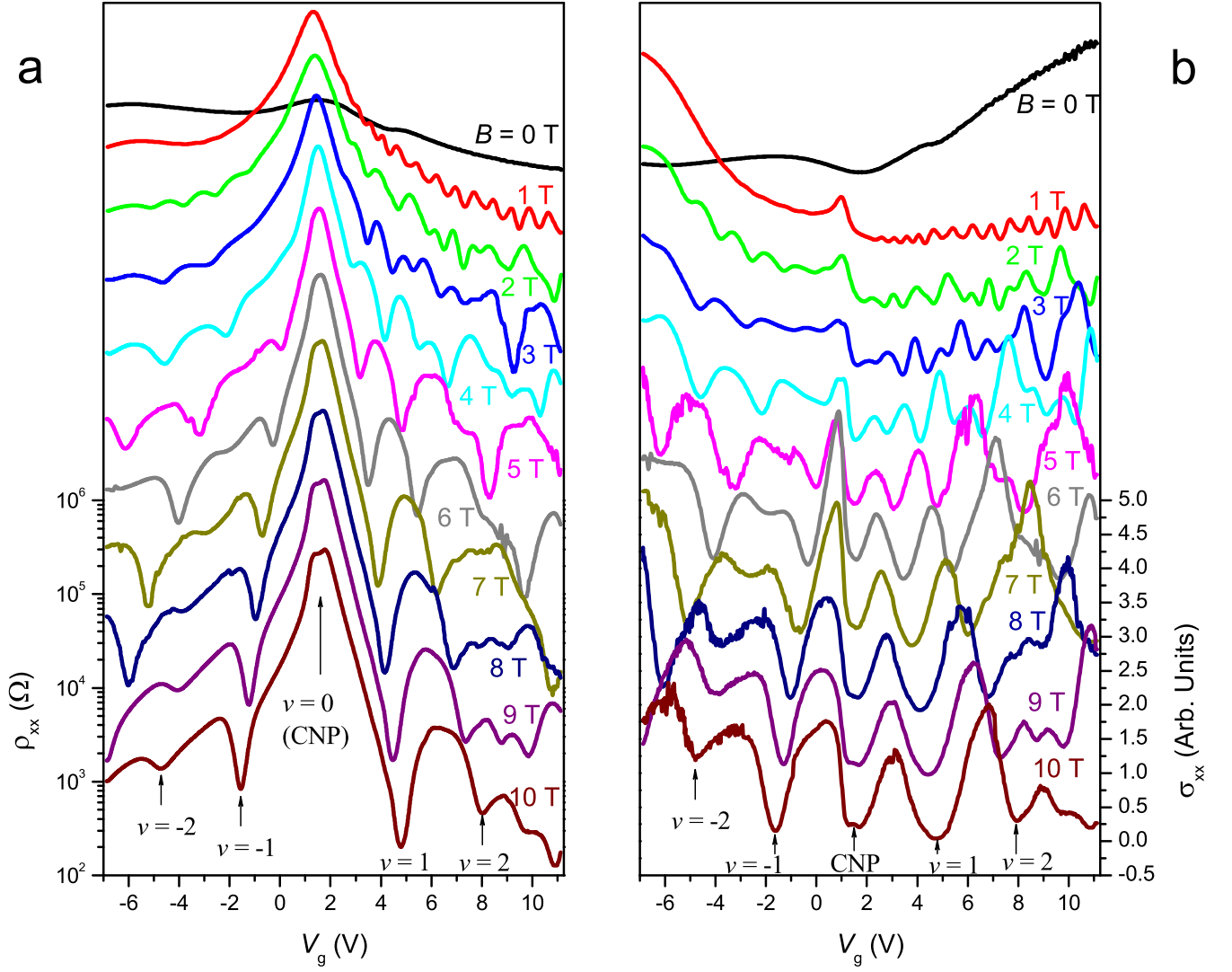


FIG. S9. Strong magnetic field magnetotransport data: **a**, $\rho_{xx}(V_g)$ dependencies at magnetic fields from 0 (top) to 10 T (bottom). Starting at $B = 9$ T each subsequent trace is shifted by 1 T on the Y-axis. **b**, $\sigma_{xx}(V_g)$ dependencies from 0 (on the top) to 10 T (bottom). Each trace was normalized by the average value $\langle \sigma_{xx}(V_g) \rangle = 1$ (see text) and shifted by 1 T on the Y-axis. The filling factors $\nu = -2, -1, 1$, and 2 are marked.

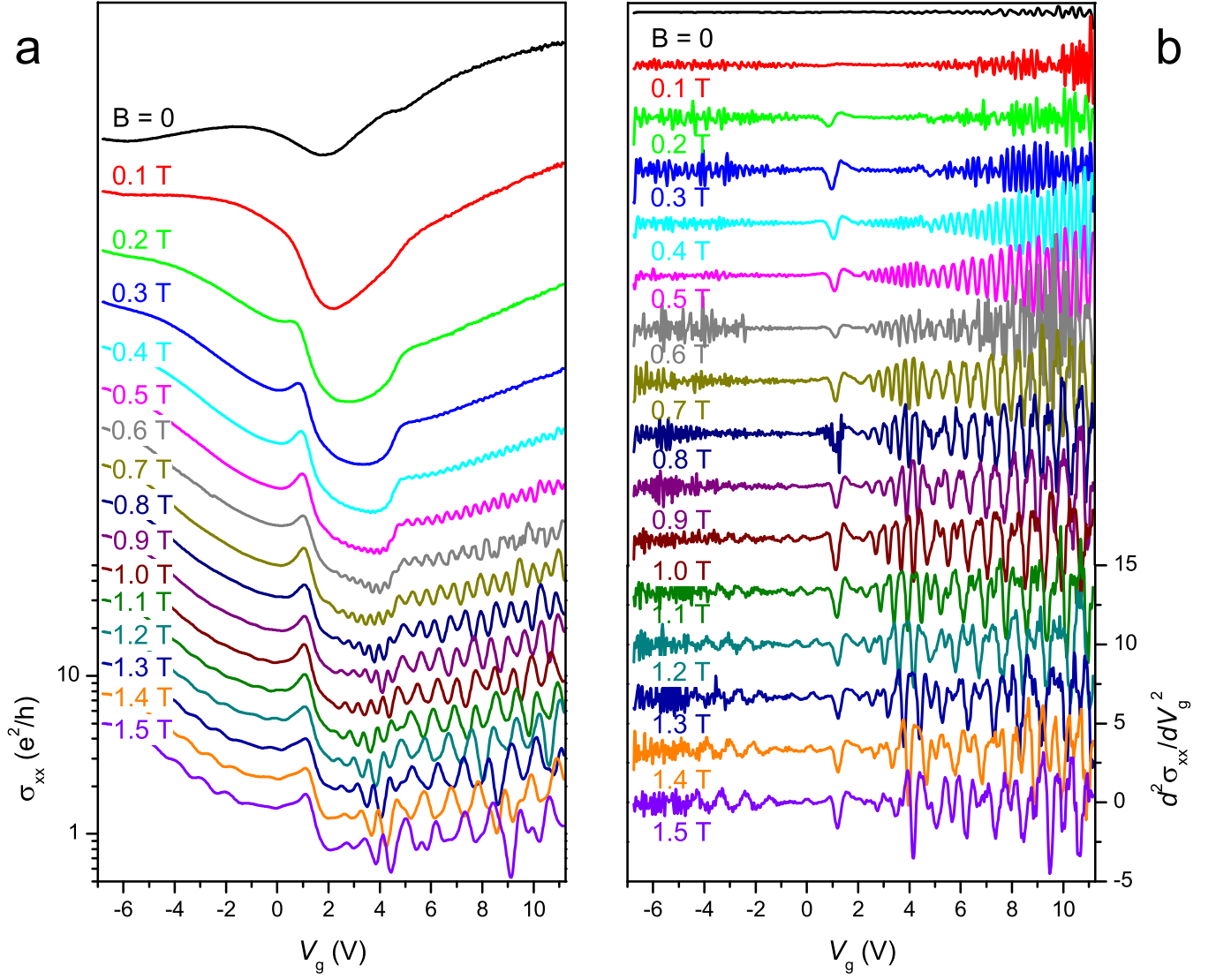


FIG. S10. Small magnetic field magnetocapacitance data: **a**, $\sigma_{xx}(V_g)$ dependencies at magnetic fields from 0 (on the top) to 1.5 T (bottom). Starting at $B = 1.4$ T each trace was shifted by 0.1 T on the Y-axis. **b** - The second derivative $d^2\sigma_{xx}(V_g)/dV_g^2$ of the traces of panel (a). The SdH oscillations come out more clearly in the second derivative. Each trace was normalized by its RMS value and shifted on the Y-axis.

-
- [1] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
 - [2] J. E. Moore, Nature **464**, 194 (2010).
 - [3] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
 - [4] Y. Ando, J. Phys. Soc. Jpn. **82**, 102001 (2013).
 - [5] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
 - [6] C. Brune, C. X. Liu, E. G. Novik, E. M. Hankiewicz, H. Buhmann, Y. L. Chen, X. L. Qi, Z. X. Shen, S. C. Zhang, and L. W. Molenkamp, Phys. Rev. Lett. **106**, 126803 (2011).
 - [7] D. A. Kozlov, Z. D. Kvon, E. B. Olshanetsky, N. N. Mikhailov, S. A. Dvoretzky, and D. Weiss, Phys. Rev. Lett. **112**, 196801 (2014).
 - [8] S. H. Groves, R. N. Brown, and C. R. Pidgeon, Phys. Rev. **161**, 779 (1967).
 - [9] K. Seeger, *Semiconductor Physics* (Springer Series in Solid State Sciences 40, 1985).
 - [10] L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).
 - [11] C. Brune, C. Thienel, M. Stuißer, J. Bottcher, H. Buhmann, E. G. Novik, C.-X. Liu, E. M. Hankiewicz, and L. W. Molenkamp, Phys. Rev. X **4**, 041045 (2014).
 - [12] O. Crauste, Y. Ohtsubo, P. Ballet, P. A. L. Delplace, D. Carpentier, C. Bouvier, T. Meunier, A. Taleb-Ibrahimi, and L. Levy, arXiv:1307.2008 (2013).
 - [13] A. M. Shuvaev, G. V. Astakhov, C. Brune, H. Buhmann, L. W. Molenkamp, and A. Pimenov, Semicond. Sci. Technol. **27**, 124004 (2012).
 - [14] A. M. Shuvaev, G. V. Astakhov, M. Muhlbauer, C. Brune, H. Buhmann, and L. W. Molenkamp, Appl. Phys. Lett. **102**, 241902 (2013).
 - [15] A. M. Shuvaev, G. V. Astakhov, G. Tkachov, C. Brune, H. Buhmann, L. W. Molenkamp, and A. Pimenov, Phys. Rev. B **87**, 121104(R) (2013).
 - [16] K.-M. Dantscher, D. A. Kozlov, P. Olbrich, C. Zoth, P. Faltermeier, M. Lindner, G. V. Budkin, S. A. Tarasenko, V. V. Belkov, Z. D. Kvon, N. N. Mikhailov, S. A. Dvoretzky, D. Weiss, B. Jenichen, and S. D. Ganichev, Phys. Rev. B / arXiv:1503.06951 **accepted**, accepted (2015).
 - [17] I. Sochnikov, L. Maier, C. A. Watson, J. R. Kirtley, C. Gould, G. Tkachov, E. M. Hankiewicz, C. Brune, H. Buhmann, L. W. Molenkamp, and K. A. Moler, Phys. Rev. Lett. **114**, 066801 (2015).
 - [18] S. Luryi, Appl. Phys. Lett. **52**, 501 (1988).
 - [19] T. P. Smith, W. I. Wang, and P. J. Stiles, Phys. Rev. B **34**, 2995 (1986).
 - [20] V. Mosser, D. Weiss, K. v Klitzing, K. Ploog, and G. Weimann, Solid State Commun. **58**, 5 (1986).
 - [21] Y. Baum, J. Bottcher, C. Brune, C. Thienel, L. W. Molenkamp, A. Stern, and E. M. Hankiewicz, Phys. Rev. B **89**, 245136 (2014).
 - [22] J. Baars and F. Sorger, Solid State Commun. **10**, 875 (1972).
 - [23] F. Stern, Appl. Phys. Lett. **43** (10), 974 (1983).
 - [24] T. P. Smith, B. B. Goldberg, P. J. Stiles, and M. Heiblum, Phys. Rev. B **32** (4), 2696 (1985).